

**PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK
MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN
MENGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada Jurusan
Informatika Fakultas Komunikasi dan Informatika**

Oleh :

ADNAN LIESTYA NUGRAHA

L 200 130 022

**PROGRAM STUDI INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

2017

HALAMAN PERSETUJUAN

**PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK
MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN
MENGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH**

PUBLIKASI ILMIAH

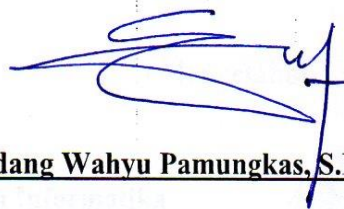
Oleh :

ADNAN LIESTYA NUGRAHA

L 200 130 022

Telah diperiksa dan disetujui untuk diuji oleh :

Dosen Pembimbing



Endang Wahyu Pamungkas, S.Kom, M.Kom.

NIK. 100.1704

HALAMAN PENGESAHAN

**PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK
MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN
MENGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH**

OLEH

ADNAN LIESTYA NUGRAHA

L 200 130 022

Telah dipertahankan di depan Dewan Penguji
Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
Pada hari Rabu, 2 Agustus 2017
Dan dinyatakan telah memenuhi syarat

Dewan Penguji :

1. Endang Wahyu Pamungkas, S.Kom., M.Kom.
(Ketua Dewan Penguji)
2. Yogiek Indra Kurniawan, S.T., M.T.
(Anggota I Dewan Penguji)
3. Diah Priyawati, S.T., M.Eng.
(Anggota II Dewan Penguji)



Mengetahui,

**Dekan
Fakultas Komunikasi dan Informatika**



Nurgiyatna, S.T, M.Sc, Ph.D.
NIK. 881

**Ketua Program Studi
Informatika**



Drs. Heru Supriyono, M.Sc
NIK. 970

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Naskah Publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 2 Agustus 2017

Penulis



ADNAN LIESTYA NUGRAHA
L 200 130 022



**UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

SURAT KETERANGAN LULUS PLAGIASI

/A.3-IL3/INF-FKI/VIII/2017

Assalamu'alaikum Wr. Wb

Biro Tugas Akhir Program Studi Informatika menerangkan bahwa :

Nama : ADNAN LIESTYA NUGRAHA
NIM : L200130022
Judul : PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK
MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN
MENGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH

Program Studi : Informatika

Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Tugas Akhir,
dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 4 Agustus 2017

Biro Tugas Akhir Informatika

Endang Wahyu Pamungkas, S.Kom., M.Kom.



UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

Feedback Studio - Google Chrome

Amah | https://www.turnitin.com/app/carta/en_us/?lang=en_us&id=18001057550000&id=855012945

turnitin

PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN MENGGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH

8 of 20

Match Overview

16%

1	eprints.ums.ac.id	Internet Source	8%
2	Submitted to Poltek... Student Paper		1%
3	Submitted to Universita... Student Paper		1%
4	Submitted to Institut B... Student Paper		1%
5	cibex2015.poli.ru.ac.id	Internet Source	1%
6	ejournal.its.ac.id	Internet Source	1%
7	Submitted to President... Student Paper		1%
8	www.ondasgudha.com	Internet Source	<1%
9	englishfirst.co.id	Internet Source	<1%
10	Submitted to Universita... Student Paper		<1%
11	pt.scribd.com	Internet Source	<1%
12	eprints.uns.ac.id	Internet Source	<1%

Page: 1 of 13 Word Count: 3238

PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN MENGGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH

Abstrak

Perkembangan komunikasi dan informasi di era ini telah didukung dengan jaringan komputer yang terdistribusi ke seluruh dunia yang dikenal dengan sebutan internet. Berbagai informasi yang diperoleh dapat berguna, salah satunya informasi dapat dimanfaatkan pihak berkepentingan dalam menunjang proses pengambilan keputusan. Informasi yang didapatkan dapat digunakan untuk mengidentifikasi kalimat dengan memanfaatkan teknik sentimen analisis. Analisis sentimen merupakan bidang studi yang menganalisis pendapat, sikap, emosi, serta evaluasi dari bahasa terungkap dalam bentuk tulisan. Analisis sentimen dilakukan dengan menggunakan metode *Lexicon Based Analysis*, yang akan memanfaatkan SentiWordnet untuk menyelesaikan masalah sentimen dengan cara mengumpulkan opini dan melakukan perhitungan nilai *sentimen*. Namun, permasalahan yang timbul dalam penggunaan metode ini adalah munculnya *ambiguitas*. Kesalahan pemilihan makna akan berakibat pada kesalahan makna kalimat secara keseluruhan. Permasalahan *ambiguitas* akan diselesaikan menggunakan algoritma similaritas *Path Length*. Hasil penelitian ini menunjukkan bahwa metode *Path Length* mampu meningkatkan performa dalam parameter *Precision* sebesar 0,5, *Recall* 0,92, dan *Accuracy* 0,61. Pemahaman terhadap konteks kalimat sangatlah penting, maka dari itu adanya *ambiguitas* harus diselesaikan untuk menemukan makna yang paling tepat.

PENDETEKSIAN AMBIGUITAS MAKNA KATA UNTUK MENINGKATKAN PERFORMA ANALISA SENTIMEN DENGAN MENGGUNAKAN ALGORITMA SIMILARITAS PATH LENGTH

Abstrak

Perkembangan komunikasi dan informasi di era ini telah didukung dengan jaringan komputer yang terdistribusi ke seluruh dunia yang dikenal dengan sebutan internet. Berbagai informasi yang diperoleh dapat berguna, salah satunya informasi dapat dimanfaatkan pihak berkepentingan dalam menunjang proses pengambilan keputusan. Informasi yang didapatkan dapat digunakan untuk mengidentifikasi kalimat dengan memanfaatkan teknik sentimen analisis. Analisis sentimen merupakan bidang studi yang menganalisis pendapat, sikap, emosi, serta evaluasi dari bahasa terungkap dalam bentuk tulisan. Analisis sentimen dilakukan dengan menggunakan metode *Lexicon Based Analysis*, yang akan memanfaatkan SentiWordnet untuk menyelesaikan masalah sentimen dengan cara mengumpulkan opini dan melakukan perhitungan nilai sentimen. Namun, permasalahan yang timbul dalam penggunaan metode ini adalah munculnya ambiguitas. Kesalahan pemilihan makna akan berakibat pada kesalahan makna kalimat secara keseluruhan. Permasalahan ambiguitas akan diselesaikan menggunakan algoritma similaritas *Path Length*. Hasil penelitian ini menunjukkan bahwa metode *Path Length* mampu meningkatkan performa dalam parameter *Precision* sebesar 0,5, *Recall* 0,92, dan *Accuracy* 0,61. Pemahaman terhadap konteks kalimat sangatlah penting, maka dari itu adanya ambiguitas harus diselesaikan untuk menemukan makna yang paling tepat.

Kata Kunci : Ambiguitas, Analisa sentimen, *Path Length*, SentiWordNet.

Abstract

The development of communication and information in this period have been supported with computer network distributed throughout the world, known as the internet. Various information obtained can be useful, the information can be utilized in teamwork to support the decision-making process. The information obtained can be used to identify the sentence by utilizing the techniques of sentiment analysis. Sentiment analysis is a field of study that analyzes the opinions, attitudes, emotions, as well as the evaluation of the language it was revealed in the form of writing. Sentiment analysis is carried out using a method of Lexicon Based Analysis, which will utilize SentiWordnet to resolve the issue of sentiment by way of gathering opinions and do calculation value of sentiment. But problems arise in the use of these methods is the appearance of ambiguity. The exact meaning of election misconduct will result in a mistake the meaning of the sentence as a whole. Problems of ambiguity will be resolved using the similaritas algorithm is the Path Length. The results of this study indicate that Path Length method can improve performance in Precision parameters of 0.5, Recall 0.92, and Accuracy 0.61. An understanding of the context of the sentence is very important, thus the existence of an ambiguity must be resolved in order to find the most appropriate meaning.

Key Words: Ambiguity, *Path Length*, Sentiment Analysis, SentiWordNet.

1. PENDAHULUAN

Perkembangan teknologi di bidang komunikasi dan informasi telah menunjukkan peningkatan yang signifikan. Perkembangan komunikasi dan informasi di era ini telah didukung dengan jaringan komputer yang terdistribusi di seluruh dunia, yang dikenal dengan sebutan internet. Terciptanya internet telah banyak berperan dalam membantu manusia dalam kehidupan sehari-hari. Tersedianya koneksi internet akan membantu manusia dalam berbagai hal. Efektif dalam menghemat waktu dan efisien dalam mengurangi biaya saat ini dapat terpenuhi dengan memaksimalkan penggunaan internet. Internet tidak hanya berperan sebagai media penyalur informasi, namun juga untuk pembelajaran jarak jauh, belanja *online*, bermain *game online*, bahkan dapat digunakan untuk media promosi dalam proses bisnis, dan sarana bergaul serta berinteraksi di media sosial.

Media sosial merupakan salah satu fasilitas untuk bersosialisasi antar individu maupun kelompok secara *online*. Media sosial berperan menciptakan sumber daya informasi dan ilmu pengetahuan di kalangan masyarakat luas. Pengguna dapat menuangkan berbagai tanggapan tentang beberapa media sosial yang telah ada. Kumpulan dari semua tanggapan subyektif yang telah dituangkan bersifat kolektif, disebut sebagai opini publik. Opini publik dapat dimanfaatkan untuk menyerap tanggapan dalam menunjang proses pengambilan keputusan tentang kepuasan pengguna, dengan menggunakan teknik analisa sentimen (Rozi, Pramono, & Dahlan, 2013).

Analisa sentimen merupakan proses mengamati, kemudian mengolahnya sesuai klasifikasi hingga terbentuk sebuah informasi yang berguna dalam pengambilan keputusan (Pang & Lee, 2008). Analisis sentimen dan pertambangan pendapat merupakan bidang studi yang menganalisis pendapat, sikap, emosi, serta evaluasi dari bahasa terungkap dalam bentuk tulisan. Penelitian tentang analisis sentimen dilakukan secara aktif dalam pembelajaran pada *data mining*, pertambangan *web*, dan pertambangan teks serta pengolahan bahasa alami (Liu, 2012). Pada penelitian Sunni & Widyanoro (2012), melakukan analisis sentimen dalam menentukan kebutuhan untuk membantu pihak berkepentingan mengetahui sentimen tentang tokoh publik yang baik atau melakukan pencitraan, dengan cara mengambil latar belakang tokoh politik. Pada dasarnya analisa sentimen merupakan klasifikasi, namun tidak sesederhana proses klasifikasi biasa karena terkait penggunaan bahasa.

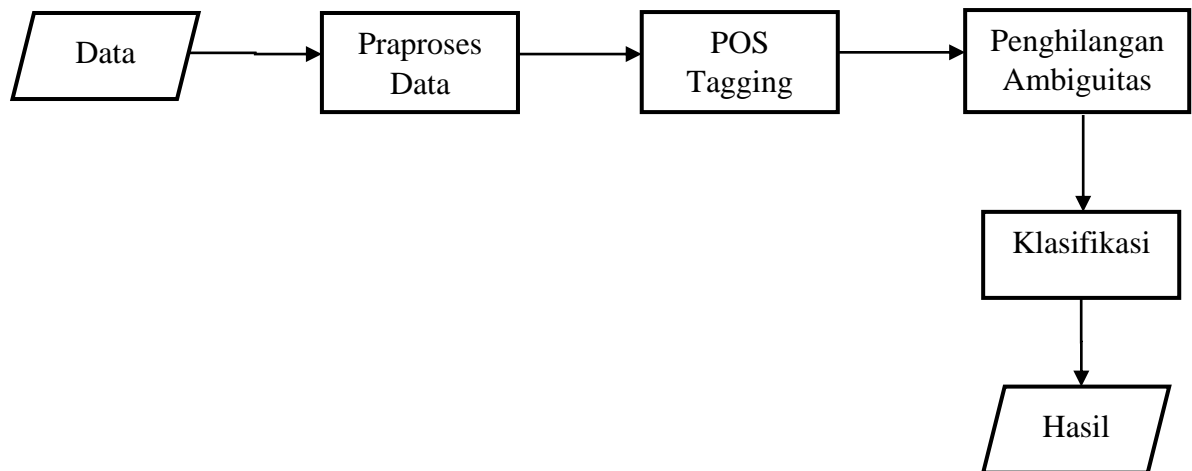
Analisa sentimen dapat dilakukan dengan dua metode pendekatan yaitu *Lexicon Based Analysis* dan *Supervised*. Metode *Lexicon Based Analysis* menyelesaikan permasalahan sentimen dengan cara mengumpulkan opini dan melakukan perhitungan nilai sentimen dengan memanfaatkan *dictionary* yang sudah sering digunakan yaitu SentiWordNet (Esuli, 2013). Kemudian metode *Supervised* merupakan pengklasifikasian teks sebagai proses untuk membentuk kelas-kelas dari dokumen berdasarkan pada kelas kelompok yang sudah diketahui sebelumnya (Darujati & Gumelar, 2012).

Penelitian ini memanfaatkan metode *Lexicon Based Analysis*. Dilakukan pengelompokan tiap kata ke dalam bentuk *synonym sets* (synset) pada struktur basis data WordNet. Kata-kata yang memiliki *sense* yang sama akan membentuk sebuah synset. Struktur WordNet memungkinkan munculnya kata dengan sintaks sama yang memiliki makna yang berbeda (*polysemous*), yang sering memunculkan ambiguitas. Ambiguitas merupakan tataran bahasa yang mempunyai beberapa semantik atau makna (Pradany & Fatichah, 2016). Kesalahan pemilihan makna akan berakibat pada kesalahan maksud kalimat secara keseluruhan, satu kata akan memiliki banyak makna dan nilai sentimen yang beda. Hal itu akan mengakibatkan terjadinya kesalahan pada klasifikasi nilai sentimen. Informasi yang tidak jelas akan sulit menghasilkan penafsiran terhadap apapun, yang akan menimbulkan keraguan.

Ambiguitas yang timbul dalam metode *Lexicon Based* dapat dihilangkan dengan algoritma similaritas *Path Length*. Namun sebelum dilakukan perhitungan *Path Length*, sebuah kalimat harus melalui proses POS Tagger terlebih dahulu, untuk memasukkan ke dalam kelas kata / *parts of speech* yang sesuai. Data yang telah dimasukkan ke dalam kelas kata yang akurat kemudian akan diproses dengan algoritma *Path Length*. *Path Length* secara sederhana menghitung jarak *node* atau relasi yang menghubungkan antar *node*. Jarak lebih pendek antara dua konsep memiliki kesamaan yang lebih tinggi (Kenett, Levi, Anaki, & Faust, 2017). WordNet akan digunakan dalam proses penghitungan *Path Length* sehingga ambiguitas dapat dihilangkan. Kata yang sudah diproses dalam WordNet akan memunculkan beberapa makna, kemudian ditentukan makna yang paling tepat sesuai dengan algoritma *Path Length*. Kemudian setelah dilakukan penghilangan ambiguitas, tahap selanjutnya yaitu pengklasifikasian sentimen dengan bantuan SentiWordNet. SentiWordNet menghitung nilai sentimen per kata. Hasil penelitian ini bertujuan untuk menghilangkan ambiguitas yang

muncul karena adanya beberapa makna, dengan algoritma similaritas *Path Length* sehingga kesalahan dalam penafsiran dapat diminimalisir.

2. METODE



Gambar 1. Alur proses penelitian

Diagram di atas merupakan alur dari metode yang diterapkan dalam penelitian ini. Data yang diambil dan dimasukkan yaitu kalimat komentar tentang pendapat pengguna media sosial, kemudian kalimat tersebut melewati tahap praproses terlebih dahulu. Kemudian dilakukan pengelompokan berdasarkan kelas kata (POS Tagger). Proses selanjutnya adalah menemukan beberapa makna melalui WordNet yang disertai penghilangan ambiguitas dengan algoritma *Path Length*. Langkah berikutnya yaitu menganalisa sentimen dengan menggunakan SentiWordNet, kemudian didapatkan nilai sentimen dari susunan kata yang membentuk kalimat. Berikut penjelasan dari alur proses pada diagram di atas secara lebih detail.

1. Praproses Data

Dilakukan normalisasi terhadap kalimat menjadi bentuk yang baku. Langkahnya sebagai berikut :

a. Penghapusan karakter

Kalimat terdapat karakter yang tidak dapat dijadikan kalimat baku, langkah yang dilakukan penghapusan terhadap karakter yang salah. Kata yang dihilangkan meliputi karakter *HTML*, kata kunci, ikon emosi, *hashtag*(#), *username*(@username), *url*(http://website.com), dan *email* (nama@website.com)

b. Penggantian kata

Sering terdapat singkatan dan ragam bahasa yang tidak resmi dalam penulisan kalimat. Harus dilakukan penggantian kata yang belum efektif sehingga akan menjadi susunan kalimat yang baku. Seperti mengganti kata “*thx*” menjadi “*thanks*”.

c. Memperpendek kata yang berlebihan

Kata-kata yang memiliki susunan huruf yang sama lebih dari dua kali diganti menjadi kata berulang yang terjadi hanya sekali. Seperti mengurangi kelebihan huruf pada kata “*nooooo*” menjadi “*no*”.

Hasil dari praproses data akan digunakan dalam proses selanjutnya yaitu POS Tagging.

2. POS Tagging

Sebelum dilakukan POS Tagging, kalimat melewati proses *tokenizing*. Proses *tokenize* dilakukan untuk mendapat daftar kata dari setiap kalimat dan memisahnya berdasarkan *stopwords*. Kemudian dilanjutkan dengan proses POS Tagging, untuk mencari *Parts of Speech* (POS) yang meliputi kata benda (*noun*), kata kerja (*verb*), kata sifat (*adjective*), dan kata keterangan (*adverb*) sesuai dengan konteks kalimat yang sesuai. Proses POS Tagging dilakukan dengan menggunakan Stanford POS Tagger. Dari dua tahap ini akan dihasilkan daftar kata penyusun kalimat, lengkap dengan *Parts of Speech*. Sebagai contoh :

“this application has always been an awesome one. i enjoy the loads of privacy it brings now.”

Hasil POS Tagger : application_NN, has_VBZ, been_VBN, awesome_JJ, enjoy_VB, loads_NNS, privacy_NN, brings_VBZ.

NN dikategorikan sebagai kata benda (*noun*), VB dikategorikan sebagai kata kerja (*verb*), JJ dikategorikan sebagai kata sifat (*adjective*).

Hasil proses ini akan digunakan pada tahap selanjutnya yaitu penghilangan ambiguitas.

3. Penghilangan Ambiguitas

Sebuah kata memiliki makna yang lebih dari satu yang dapat menyebabkan ambiguitas. Kemudian untuk mencari tahu arti kata yang paling tepat dalam kalimat tertentu dilakukan proses *Word Sense Disambiguation* (WSD) dengan algoritma *Path Length*.

Contoh kata yang dapat bernilai positif atau negatif yang mempengaruhi perhitungan nilai sentimen total, sebagai berikut.

“it's simple and very easy to use”

Dalam perhitungan dengan metode *path length* hasilnya bernilai positif, yaitu:

POS Tagger = 's_VBZ, simple_JJ, easy_JJ, use_VB

Tabel 1. Hasil perhitungan sentimen

ID synset	Positif	Negatif
02174896 (simple#1)	0.125	0.375
00749230 (easy#)	0.625	0.25
01158872 (use#1)	0	0
Skor sentimen total	0.125 (Positif)	

Kata “easy” dengan POS *adjective* dalam SentiWordnet memiliki akumulasi skor positif dan skor negatif. Salah satu synset “easy” dengan ID 00749230 memiliki skor positif 0,625 dan skor negatif 0,25 sehingga menghasilkan total skor positif (0,125). Synset “easy” dengan ID 01272176 memiliki skor positif 0 dan skor negatif 0,625 sehingga menghasilkan total skor negatif. Pada contoh kalimat diatas, metode *path length* memilih synset “easy” dengan ID 00749230 yang total skor positifnya menghasilkan total nilai sentimen positif (0,125) sehingga sesuai dengan opini yang ditentukan. Jika dipilih synset “easy” dengan ID 01272176 maka total nilai sentimen pada kalimat tersebut adalah negatif (-0,875), tentu hasil itu tidak sesuai dengan opini yang sudah ditentukan.

Path Length digunakan untuk mengukur jarak *node* atau relasi yang menghubungkan antar *node*. Untuk mengukur kesamaan semantik antara dua synsets (*synonym sets*), menggunakan *hyponym/hypernym* (hubungan). Proses pemilihan synset dipilih berdasarkan *Part of Speech* dari kalimat. Cara sederhana untuk mengukur kesamaan semantik antara dua synsets adalah untuk menghilangkan taksonomi (pengelompokan) dan mengukur jarak antar synset di WordNet. WordNet berisi informasi tentang kata benda, kata kerja, kata sifat dan kata keterangan. WordNet mengorganisir konsep yang terkait ke dalam kumpulan sinonim atau synset. Masing-masing synset mewakili suatu konsep atau makna / pengertian kata. Menghitung *Path Length* dengan rumus berikut :

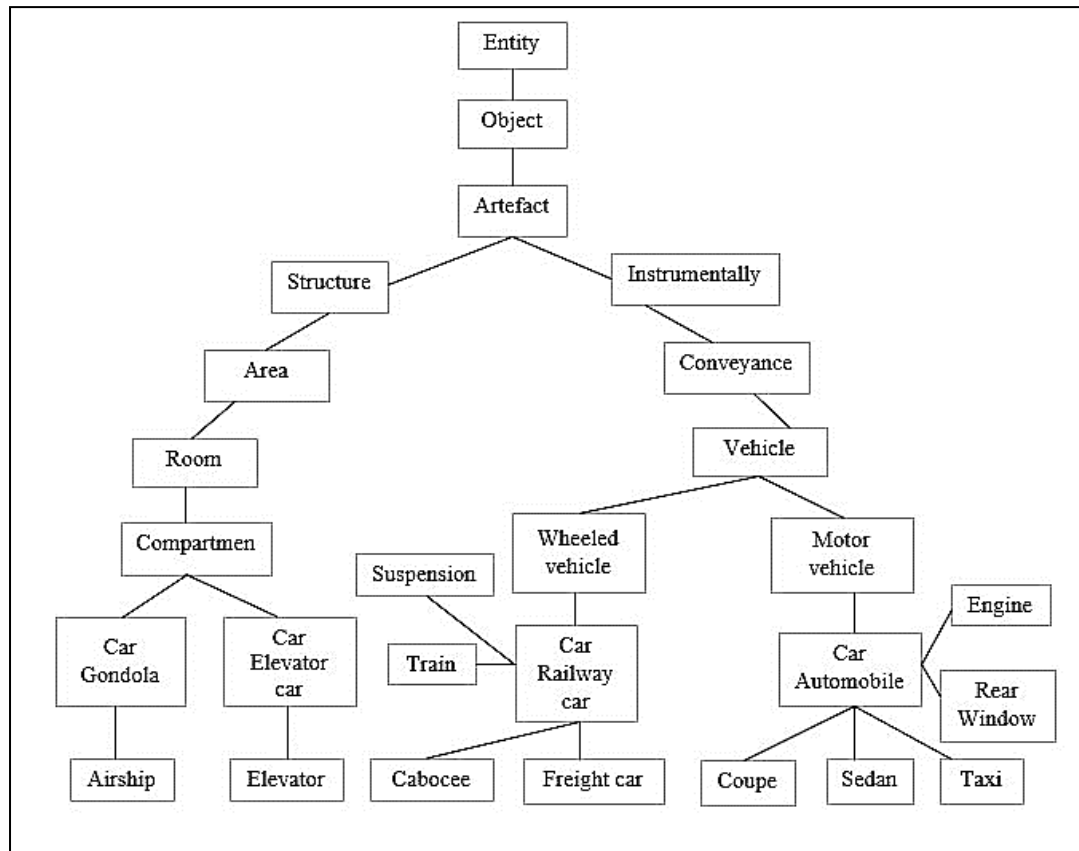
$$\text{Sim}(s, t) = 1 / \text{distance}(s, t).$$

Keterangan :

s dan t : menunjukkan sumber dan target kata-kata yang dibandingkan.

distance adalah panjang jalur dari s ke t menggunakan simpul perhitungan.

Contoh perhitungan *Path Length* sebagai berikut :



Gambar 2. Pohon akar WordNet

Dalam struktur pohon akar di WordNet, untuk menghitung antara synset *airship* dan *room*, maka rumusnya adalah $1/\text{distance}$. Jarak antara *airship* dan *room* adalah 4, maka hasil *path length* adalah $1/4$ atau 0,25. Contoh lain, antara synset *room* dan *motor vehicle*, jarak antar keduanya adalah 8, maka hasilnya adalah $1/8$ atau 0,125.

Pada metode ini akan dipilih synset yang paling tepat, caranya dengan memilih hasil *path length* yang paling besar atau jarak synset yang terpendek. Hasil serangkaian proses ini akan digunakan pada tahap selanjutnya yaitu klasifikasi.

4. Klasifikasi

Dilakukan pengklasifikasian sentimen per kata dan menghitung nilai sentimen total dalam sebuah kalimat. Perhitungan sentimen per kata menggunakan *lexicon* SentiWordNet. Langkah berikutnya yaitu menentukan nilai sentimen dari keseluruhan konteks kalimat, dengan rumus sebagai berikut.

$$S_{positive} = \sum_{i \in t}^n positive\ score_i \quad (1)$$

$$S_{negative} = \sum_{i \in t}^n negative\ score_i \quad (2)$$

Kedua persamaan di atas adalah rumus untuk menghitung orientasi semantik dari suatu kalimat. Dari masing-masing suku kata dalam sebuah kalimat, semua skor positif dan skor negatif akan dijumlahkan secara terpisah. Kemudian menentukan orientasi semantik dengan rumus sebagai berikut.

$$Sentence_{sentiment} \begin{cases} positive\ if\ S_{positive} > S_{negative} \\ neutral\ if\ S_{positive} = S_{negative} \\ negative\ if\ S_{positive} < S_{negative} \end{cases} \quad (3)$$

Persamaan di atas menunjukkan bahwa dengan cara membandingkan jumlah skor positif dan skor negatif akan menentukan sentimen dari kalimat. Jika skor positif lebih besar dari skor negatif, maka hasilnya positif. Jika skor positif kurang dari skor negatif, maka hasilnya negatif. Dan jika skor positif sama dengan skor negatif, maka hasilnya netral (Pamungkas & Putri, 2016).

3. HASIL DAN PEMBAHASAN

Data yang digunakan didapat dari opini pengguna beberapa aplikasi *chatting* yaitu BBM, Line, dan Whatsapp. Didapatkan 334 data, diantaranya terdiri dari 131 opini positif dan 203 opini negatif, yang telah diklasifikasi secara manual.

Peningkatan performa analisa sentimen dapat dibuktikan dengan cara membandingkan teknik *Lexicon Based* menggunakan metode *Path Length* dengan metode *First Sense* dan metode *Path Length* dengan metode *Average*. Pada penelitian sebelumnya menggunakan metode *First Sense* hasilnya dirasa kurang baik, karena terjadi cukup banyak ketidaksesuaian antara klasifikasi sentimen manual dan hasil dari sistem (Kusumawati & Pamungkas, 2017). Peningkatan performa tersebut dapat dilihat dalam Tabel 2.

Tabel 2. Perbandingan hasil pengujian

	Metode		
Parameter	Path Length	First Sense	Average
Precision	0,5	0,41	0,45
Recall	0,92	0,67	0,65
Accuracy	0,61	0,53	0,55

Tabel diatas merupakan hasil perbandingan pengujian algoritma *path length* dengan metode *first sense* dan *average*. Pengujian ini dilakukan untuk mengukur peningkatan performa sistem dalam menganalisa sentimen yang dikategorikan dalam parameter *precision*, *recall* dan *accuracy*. *Precision* adalah tingkat keakuratan antara data klasifikasi secara manual dan keluaran yang dihasilkan oleh sistem. *Recall* adalah tingkat kesuksesan sistem saat mendapatkan kembali data yang diuji. *Accuracy* adalah tingkat kedekatan antara perkiraan dengan nilai yang sesungguhnya.

Hasil perbandingan menunjukkan selisih yang signifikan dalam nilai *recall* antara metode *Path Length* dan dua metode *Lexicon Based* dengan *First Sense* dan *Average*, yaitu 0,67 (*First Sense*), 0,65 (*Average*), dan 0,92 pada metode *Path Length*. Nilai *precision* dan *accuracy* juga menunjukkan peningkatan, namun tidak signifikan.

Nilai *precision* mengalami peningkatan 0,09 dan 0,05. Nilai *accuracy* juga mengalami peningkatan sebesar 0,08 dan 0,06. Nilai *precision* dan *accuracy* mengalami peningkatan yang hampir sama. Nilai *recall* menunjukkan peningkatan yang paling besar, ini membuktikan bahwa algoritma similaritas *Path Length* mampu meningkatkan performa sistem dalam mendapatkan kembali data yang sedang diuji. Peningkatan nilai *precision* dan *accuracy* yang rendah dipengaruhi oleh beberapa data uji yang beropini negatif namun oleh sistem menghasilkan keluaran positif, dan sebaliknya. Salah satu contoh terdapat pada kalimat “*have many annoying notifications of useless animations and stickers, it is better to remove all stickers. video calls are slow with interrupted streaming*”, kalimat tersebut beropini negatif, dan hasil perhitungan oleh sistem dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 3. Proses POS Tagging

Word	POS
Have	VB
Many	JJ
Annoying	JJ
Notifications	NNS
Useless	JJ
Animations	NNS
Stickers	NNS
Is	VBZ
Better	JJR
Remove	VB
Stickers	NNS
Video	NN
Calls	NNS
Are	VBP
Slow	JJ
Interrupted	JJ
Streaming	NN

Tabel 4. Hasil sentimen

ID synset	Skor Positif	Skor Negatif
02203362 (have)	0.25	0
01551633 (many)	0	0
00089550 (annoying)	0	0.5
02497141 (useless)	0.125	0.625
00230335 (better)	0.875	0
00173338 (remove)	0	0
06277803 (video)	0	0
00980527 (slow)	0	0

00598413 (interrupted)	0	0
11443929 (streaming)	0	0
Hasil sentimen	1.25	1.125
Tipe data	Positif (0.125)	

Dari perhitungan diatas menunjukkan bahwa data yang diproses menghasilkan sentimen positif, sedangkan asumsi yang dilakukan secara manual dikategorikan dalam opini negatif. Terdapat beberapa permasalahan yang menyebabkan hasil analisa menjadi kurang maksimal, sebagai berikut.

1. Data tidak seimbang

Data yang digunakan didapat dari opini pengguna beberapa aplikasi *chatting* yaitu BBM, Line, dan Whatsapp. Didapatkan 334 data, diantaranya terdiri dari 131 opini positif dan 203 opini negatif, yang telah diklasifikasi secara manual. Data opini yang terpaut selisih yang cukup banyak mengakibatkan keterbatasan dalam pengambilan salah satu opini, yaitu opini positif yang hanya 131 data. Opini positif yang terpaut selisih 72 data dengan opini negatif akan mempengaruhi peningkatan nilai *precision*, yang hanya sekitar 0,05 sampai 0,09. Jika data opini positif seimbang dengan data opini negatif maka kemungkinan besar peningkatan nilai *precision* juga semakin besar.

2. Kesalahan pendeteksian opini

Kondisi ini terjadi karena pada data opini yang tergolong positif terdeteksi oleh sistem sebagai nilai sentimen negatif. Terdapatnya kata “*other*” akan mempengaruhi hasil analisa yang dilakukan oleh sistem, sebagai contoh pada opini “*if there is no BBM then there is no any other messenger. i love it, it's second to none*”. Kalimat tersebut tergolong kedalam opini positif, tetapi sistem mendeteksinya sebagai sentimen negatif, dengan total skor sentimen -0,125. Keluaran itu terjadi karena sistem memilih synset “*other*” dengan ID 02069355 yang total skornya negatif (0,625). Jika sistem memilih synset “*other*” dengan ID 00969333 yang total skornya netral (positif 0,375 negatif 0,375) maka hasil total skor sentimennya menjadi 0,5 yang artinya kalimat tersebut tergolong sebagai sentimen positif.

4. PENUTUP

Penelitian ini bertujuan untuk menghilangkan ambiguitas yang muncul karena adanya beberapa makna, dengan algoritma similaritas *Path Length* sehingga kesalahan dalam penafsiran dapat diminimalisir. Berdasarkan pembahasan yang sudah dikemukakan diatas maka dapat ditarik kesimpulan sebagai berikut.

1. Penggunaan metode *Path Length* menghasilkan performansi lebih baik dibanding metode *First Sense* dan *Average*. Hal itu dapat dilihat dari hasil *precision*, *recall* dan *accuracy* yang mengalami peningkatan.
2. Masalah ambiguitas dapat teratasi karena sistem mampu menentukan synset yang tepat dari beberapa kata yang dimasukkan, namun tidak semua hasil pengujian tiap kalimat sesuai dengan opini yang sebenarnya.
3. Dalam mengukur performansi dalam kategorisasi, hasil *precision*, *recall* dan *accuracy* akan dipengaruhi oleh jumlah data yang digunakan. Hasil klasifikasi akan semakin baik jika semakin banyak jumlah data yang digunakan.
4. Data yang tidak seimbang antara opini positif dan opini negatif akan mengakibatkan hasil performansi dengan metode *Path Length* kurang maksimal, untuk itu diharapkan digunakannya data uji yang seimbang dalam penelitian selanjutnya.
5. Proses *Word Sense Disambiguation* dengan algoritma *Path Length* masih menemukan kendala yaitu kesalahan dalam pendeteksian opini karena pemilihan synset yang kurang tepat oleh sistem. Hal ini mungkin dapat dipecahkan oleh proses WSD dengan algoritma lain, salah satunya yaitu *Wu & Palmer*.

DAFTAR PUSTAKA

- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia. *Jurnal Bandung Text Mining*, 16(1), 5-1.
- Esuli, A. (2013). The user feedback on sentiwordnet. *arXiv preprint arXiv:1306.1343*.
- Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of experimental psychology. Learning, memory, and cognition*.

- Kusumawati, I., & Pamungkas, E. W. (2017). Analisa Sentimen Menggunakan Lexicon Based Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Rokok Pada Media Sosial Twitter (*Doctoral dissertation, Universitas Muhammadiyah Surakarta*).
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Pamungkas, E. W., & Putri, D. G. P. (2016, August). An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia. In *Engineering Seminar (InAES), International Annual* (pp. 28-31). IEEE.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Pradany, L. N., & Fatichah, C. (2016). Analisa Sentimen Kebijakan Pemerintah Pada Konten Twitter Berbahasa Indonesia Menggunakan SVM dan K-Medoid Clustering. *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, 11(1), 59-66.
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2013). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37-43.
- Sunni, I., & Widyantoro, D. H. (2012). Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini terhadap Tokoh Publik. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, 1(2).